Subject Matter Expert (SME)/Peer Use in M&S V&V

Dale K. Pace, *Principal Professional Staff*
Johns Hopkins University Applied Physics Laboratory (JHU/APL)
11100 Johns Hopkins Road
Laurel, Maryland 20723-6099
(240) 228-5650; (240) 228-5910 (FAX)
dale.pace@jhuapl.edu

Jack Sheehan, *Program Manager for Knowledge Integration*
C3I & Strategic Systems Directorate
Director of Operational Test & Evaluation (DOTE)
4850 Mark Center Drive #10606
Alexandria, Virginia
(703) 681-1440 ext. 110
Jack.Sheehan@osd.mil

Defense Modeling and Simulation Office (DMSO)
1901 N. Beauregard #500
Alexandria, Virginia
(703) 998-0660 x448
jsheehan@dmso.mil

The Invited Paper for Session A4 of the Foundations for V&V in the 21st Century Workshop (Foundations '02) held at the Johns Hopkins University Applied Physics Laboratory Laurel, Maryland (USA) October 22-24, 2002.

Foundations '02 Workshop Sponsors (listed alphabetically; bold = initiating sponsor)

- Association for Computational Machinery (ACM) Transactions on Modeling and Computer Simulation (TOMACS)
- Aegis Technology Group, Inc.
- American Society of Mechanical Engineers (ASME) Applied Mechanics Division (AMD)
- Arizona Center for Integrative Modeling & Simulation (ACIMS)
- Boeing Phantom Works
- Center for Economic Research (CentER), Tilburg University (Netherlands)
- Clemson University
- **Defense Modeling and Simulation Office (DMSO)**
- Federal Aviation Administration (FAA) Flight Standards Service
- Gesellschaft für Informatik (Bonn, Germany)
- Illgen Simulation Technologies
- Innovative Managament Concepts, Inc. (IMC)
- Johns Hopkins University/Applied Physics Laboratory (JHU/APL)
- Joint Accreditation Support Activity (JASA)
- **Modeling and Simulation Subcommittee of the Joint Army Navy NASA Air Force (JANNAF) Interagency Propulsion Committee**
- McLeod Institute of Simulation Sciences -- California State University (CSU), Chico
- Ministry of Defence (UK) Synthetic Environment Coordination Office
- Modeling & Simulation Information Analysis Center (MSIAC)
- NASA Ames Research Center
- National Institute of Standards and Technology (NIST)
- *National Training Systems Association (NTSA)* – hosting sponsor
- Office of Naval Research (ONR)
- Shodor Education Foundation, Inc.
- Simulation Interoperability Standards Organization (SISO)
- Survivability/Vulnerability Information Analysis Center (SURVIAC)
- The Society for Modeling and Simulation International (SCS)
- United States Association for Computational Mechanics (USACM)
- University of Central Florida Institute for Simulation and Training (UCF/IST)

Table of Contents

Table of Contents (continued)

Abstract

Review by experts and peers is a primary validation technique in all aspects of software and system development as well as for models and simulation. This kind of validation technique is widely used both because of its capabilities to provide insight about model and simulation capabilities and because adequate real world (experimental) data often are not available to allow robust quantitative validation. Ways that reviewers are selected, managed, and used are very important, and so are how reviews are conducted and responses to them. Review use of experts and peers is an area in which there is very little documented guidance among modeling and simulation, software, and systems engineering communities -- no formal standards for such reviews exist in any of these communities. This paper a) identifies the guidance that exists about reviews in modeling and simulation, b) describes current practices with identification of strengths and weaknesses, c) examines processes used within the legal system to determine expert witness qualifications and extracts useful principles applicable to modeling and simulation, and d) suggests both research endeavors and best practices that can lead to improvements in modeling and simulation reviews.

1. Introduction

This introduction has four parts. First, we provide a clear description of the subject that we address, one that defines its scope and boundaries. Second, we identify the primary pertinent literature so that the reader has a starting point in reviewing the literature. By primary pertinent literature, we mean the handful of sources that one should turn to in order to gain a substantive understanding of this topic. Third we give you a bit of information about ourselves so that you can judge our comments in light of our backgrounds. Finally we provide information about the scope and structure of the paper so that you will know what to expect in it.

1A. Delineation of the Topic. This paper is about reviews of models and simulation, especially reviews that are related to validation assessments. The primary kind of review that we address are reviews by experts. Sometimes expert review is called "peer review." Sometimes the experts are called "subject matter experts" or SMEs. Sometimes a validation review of simulation results by experts is called "face validation." And sometimes the reviews are just called reviews. This paper will address all aspects of such reviews: the variety of review purposes, how reviewers are selected, oriented (or trained), and managed; how reviews should be conducted and documented – this includes knowledge engineering processes related to obtaining and organizing information from the experts; limits on capabilities of reviews; and legal and professional aspects of reviews. In addition, the paper addresses current research and research needs relative to modeling and simulation reviews.

1B. Primary Pertinent Literature Relative to Modeling and Simulation Review. Full bibliographic information for all of the resources identified here may be found in the Reference section at the end of this paper; appropriate elements of that material also has been included in the Foundations '02 verification and validation (V&V) bibliography. We also provide Universal Resource Locator (URL) information for many of the resources. All of these URLs worked in the spring of 2002 (mainly May or June), but, as is well known, URLs often cease to work when websites are changed, materials are removed or relocated, etc. Therefore, we cannot promise that the URLs will work when a reader tries to obtain the material on a site.

DoD RPG. The best general guidance for subject matter expert (SME[1]) use in modeling and simulation (M&S) verification, validation, and accreditation (VV&A) may be found in the special topic on SME use in the Department of Defense (DoD) Recommended Practices Guide (RPG) for M&S VV&A, Millennium edition, 2000. It addresses all aspects of SME selection, use, and management; it also identifies problems that often occur when SMEs are used in M&S VV&A. *Subject Matter Experts and VV&A*, http://www.msiac.dmso.mil/vva/default.htm. Large portions of the RPG SME material appear nearly verbatim in this paper.

Expert Qualifications & Testimony. A 1999 synopsis of legal rulings about who qualifies as an expert and how expert testimony may be used in legal proceedings. Many valuable parallels exist for guiding expert use in M&S VV&A. Paul C. Giannelli, Albert J. Weatherhead III, and Richard W. Weatherhead, "Expert Qualifications & Testimony," *Conference on Science and Law*, San Diegp, April 15, 1999, http://www.scientific.org/distribution/law-review/giannelli.pdf.

SME Quality Control. A current endeavor to determine if formal processes can bring enough stability and quality assurance to evaluations by SMEs to allow their expectations to meaningfully serve as a surrogate for data in validation of simulation results for which real-world data are very limited or not available. Michael Metz and Scott Harmon, "Using Subject Matter Experts for Results Validation – A Progress Report," paper 02S-SIM-095 at *Spring 2002 Simulation Interoperability Workshop (SIW),* Orlando, FL, April 2002, http://www.sisostds.org/conference/View_Public_Number.cfm?Phase_ID=2.

1C. Authors Backgrounds.

Dale K. Pace became concerned about simulation correctness and credibility in the 1960s, and has been concerned about such for simulations used at his own organization since then. He has led teams performing V&V reviews on simulations created and/or used by organizations other than his own since the early 1980s. Those V&V reviews often involved SMEs. He has selected SMEs, trained and oriented them, established review processes, managed SMEs, and merged information from SMEs with other V&V evidence to create realistic appreciation for simulation capabilities and limitations. He has published a number of items about SME use: a mid-1980s paper describing the factors that made a cross-department (DoD/DOE) simulation review successful [Pace 1986], a section on SME use in a 1998 M&S text [Pace 1998], papers at conferences such as the Simulation Interoperability Workshop (SIW) [Pace 1999], and contributions to the Millennium Edition (2002) of the *RPG* special topic on SME use. Dr. Pace is a member of the Principal Professional Staff of the Johns Hopkins University Applied Physics

[1] Most acronyms have more than one possible referent. This is certainly true for "SME". In Asian and European commercial (and academic) arenas, SME is an acronym for "small and medium enterprise" or "small manufacturing enterprise," which is why searches about managing SMEs will lead one to courses in business administration. Many references to "software SMEs" also have a organizational reference for small and medium size enterprise instead of an expert connotation.

Laboratory, taught in the graduate technical management program of Hopkins' Whiting School of Engineering for a decade, and is a Program Co-chair for Foundations '02.

Jack Sheehan is currently assigned to an Interagency Personnel Act (IPA) position in the Office, Director of Operational Test and Evaluation (DOT&E), Strategic and Command, Control, Communications and Intelligence (C3I) Systems as the Modeling and Simulation (M&S) liaison. He also serves as the Knowledge Integration (KI) Program Manager and Technical Advisor to the Director, Defense Modeling and Simulation Office (DMSO). Assigned as an IPA to DOT&E in September, 2001 and to DMSO since January 1996, Mr. Sheehan manages programs requiring both problem-domain experience and implementation-domain expertise to create and deploy solutions for the warfighter. Mr. Sheehan has concentrated the problem of unambiguously communicating warfighter requirements to systems engineers methods to include computer-based systems interfaces, executable mission content, data exchange, interoperability logical information content and formal program interfaces. Mr. Sheehan is affiliated with the Applied Research Laboratory, University of Texas at Austin (ARL:UT) and was the senior systems engineer conducting basic research, architecture definition, engineering design, prototype implementation, and at-sea feasibility demonstration for US Navy combat systems from August 1985 to January 1996.

1D. Scope and Structure. This paper begins with **descriptive definitions** of terms that may be applied to personnel used in modeling and simulation reviews, especially reviews that may be related to any aspect of M&S VV&A. In that context, we discuss possible **roles** and **activities** for such personnel. In general, we shall refer to these people as SMEs. Whenever we need to make a significant functional, legal, or operational distinction, such as noting the different connotations attached to SME, peer review, and expert witness, we will take great pains to ensure that such distinctions are clear. In all other circumstances, as a matter of linguistic convenience, we will simply use the term "SME" to refer to all varieties of review personnel. Next we address processes for **selecting** and **using** SMEs. Then we identify **problems** often associated with SMEs and suggest how they can be avoided or mitigated. This leads us to **issues** associated with SMEs in M&S VV&A, and facilitates identification of **research needs**. We end with **recommendations**, **conclusions,** and **comments**. Knowledge engineering principles and practices are addressed mainly in our discussion of SME activities; administrative and management considerations appear primarily in sections on SME selection and use.

2. Key V&V-related M&S Review Processes, Techniques, and Technologies.

2A. Descriptive Definitions. We begin by describing who is expected to perform M&S V&V-related reviews. We address three groups: 1) peers, as in a "peer review," 2) technical experts, and 3) SMEs. We will take these in reverse order.

We will use the SME description/definition in the *RPG* [DoD 2000].

> **Subject Matter Expert (SME):** An individual who, by virtue of position, education, training, or experience, is expected to have greater-than-normal expertise or insight relative to a particular technical or operational discipline, system, or process, and who has

been selected or appointed to participate in development, verification, validation, accreditation or use of a model or simulation.

The RPG notes that this definition is compatible with "expert" in legal parlance:

> **Technical Expert** (in legal parlance)**:** "an individual who, by virtue of his or her specialized knowledge and experience, can explain, through competent testimony, a technical matter that lies outside the understanding of the average lay person . . . An expert may base his opinion on facts and documents not in evidence, as long as those facts and documents are reasonably relied upon by experts in his field" [Friedman & Kremen, 1997]. "The factors the Supreme Court listed as relevant to the inquiry [for use in assessing the admissibility of scientific expert testimony] are: 1) whether the expert's technique can be tested through the scientific method; 2) whether the technique was subject to peer review and publication; 3) the known or potential rate of error of the technique; 4) the existence and maintenance of standards controlling the technique's operation; and 5) whether the technique had gained general acceptance in the relevant scientific community." [New Jersey Judiciary 2002].

It should be noted that even though there are strict guidelines, as indicated above, about who may testify in court as an expert, such guidelines do not ensure unanimity of views, as is well-known from the many court cases in which conflicting views have been presented by experts who satisfy court criteria to testify as expert witnesses. This leads one to expect that there is the possibility of similar diversity in conclusions from those reviewing M&S.

Peer review is a standard part of academic life. For many academics, one must publish in referred (i.e., peer reviewed) journals or stagnate professionally. In that context, the "peer" is one who possesses the knowledge and competence to evaluate an article for factual and process soundness, and to know if something proposed for publication already exists in the literature. The internet and the often long time lag for print publication have posed a serious challenge to traditional peer review processes for academics [Roberts 1999]. Peer review is also a standard part of engineering, and has also become part of software engineering -- e.g., "Peer Reviews" are expected as a regular part of software development in organizations that qualify for higher Software Engineering Institute (SEI) Capability Maturity Model (CMM) ratings [SW-CMM 1997]. There is another aspect to peer review. It implies relative independence on the part of the reviewer so that candor is more likely than might occur if the reviewer is dependent upon those whose material is being reviewed. This brings us to the following definition:

> **Peer Review:** review by one (or more) who possess enough independence from those responsible for the item to be reviewed that candor is likely and who possess adequate knowledge and skill to evaluate the item so that flaws (if present) are likely to be detected and general capabilities and limitations of the item can be understood and appreciated (i. e., put into a proper context and perspective).

It should now be clear why we have taken time to describe and define SME, technical expert, and peer review. These review processes overlap and have much in common, but there are differences, which at times can be important. Peer review stresses both competence and independence; technical experts need both competence and communication ability (to satisfy legal requirements); and SMEs can be qualified because of vested interests, not just competence. However, as noted previously, we shall loosely use the term "SME" for reviews by any of these three in M&S VV&A, unless we need to emphasis a distinctive aspect implied by one of the term's proper definition. Likewise, we shall use the term "expertise," whether the "expert knowledge" that is implied by this term is technical, application domain related, or a consequence of one's vested interests.

2B. Possible SME Roles. As noted in the RPG, SMEs can have many roles in M&S development and use. Some of these specifically pertain to VV&A, and others do not. We discuss possible SME roles in three categories: 1) application domain expertise; 2) simulation design, development, or implementation expertise; and 3) VV&A expertise.

Abstraction is a mental facility that permits humans to view real world problems with varying degrees of detail depending on the current context of the problem. Indeed the use of abstraction may be the most "human" of all homo sapien activities. Models and simulations are the embodiment of such abstractions. Easily the most important function that an SME performs in VV&A is to provide a technical judgment, based on objective theory and professional expertise, on the degree to which the abstraction implemented in a particular model or simulation is suitable for the purpose for which that model or simulation is to be employed. An abstraction which is entirely suitable for one purpose may be manifestly unsuitable for another.

The Functional Descriptions of the Mission Space (FDMS) attempts to capture (and formally document) the abstractions that warfighters employ to conceive and conduct real military operations. The conceptual model in the VV&A RPG is an attempt to capture (and formally document) the abstractions that the M&S developer intended to implement. It is the essential role of the SME to determine the abstraction that is actually embodied in the M&S and to determine the match (or mismatch) of the actual abstraction with the purpose at hand.

2B1. **Application Domain Expertise Roles.** When simulation development begins (and sometimes before it begins), domain expertise SMEs are needed to create an authoritative description of the application domain which the simulation is to address. Once simulation objectives have been established and stated as requirements for the simulation, development of the simulation conceptual model may begin, although sometimes conceptual model development will occur in parallel with development of M&S requirements. Normally, the first step in conceptual model development[2] for the simulation is to collect authoritative information about the intended application domain since that information provides constraints on viable concepts for a simulation that can satisfy its requirements. However, development of the conceptual model and collection of authoritative information about the application domain have enough "chicken-egg" intertwining characteristics that either can come first.

Application domain SMEs collect and organize information about the application domain that is pertinent and appropriate from the perspective of intended use of the simulation. For simulations

---

[2] In the *RPG*, the simulation conceptual model is how the simulation developer transforms simulation requirements into specifications with enough detail to ensure that a simulation designed and implemented according to them will fully satisfy simulation requirements.

with representations of physical systems and processes, information about the application domain includes the laws of physics, chemistry, biology, etc., facts and observations from tests and experiments, and engineering principles. For simulations of military activities, such SMEs may be military personnel with adequate operational experience and insight to know how the military operates. The Defense Modeling and Simulation Office (DMSO) has engaged in a knowledge engineering endeavor to collect and organize such information, possibly facilitating its reuse. A variety of knowledge acquisition tools (KATs)[3] and processes were developed to facilitate this endeavor, which is now known as the Functional Description of the Mission Space (FDMS).[4] SMEs are used to provide the application domain knowledge, to aid in capturing, organizing, maintaining, and disseminating that knowledge, and to facilitate its utility in military simulations.

The FDMS provides the simulation developer with support for Functional Description creation, integration, and maintenance within DoD simulation programs, and interoperability across DoD simulation programs through integration and interoperability standards, Common Semantics and Syntax (CSS), Data Interchange Formats (DIFs), a closed-loop engineering process, and an operational infrastructure. These have been the means to extract knowledge about the application domain from experts (military operational personnel) who are not well versed in simulation development or formal methods of describing their knowledge. FDMS provides simulation implementation-independent descriptions of real world processes, entities, and environment that facilitate V&V of application domain knowledge representations and their potential reuse.

Processes have also been developed to ensure that the knowledge engineer has captured insights from a domain expert correctly. Natural Language Modeling (NLM) is one such method, whose applications can employ mathematically precise information analysis procedures developed by John Sharp [1998].

2B2. Simulation Design, Development, & Implementation Expertise SMEs.

SMEs having computer hardware or software expertise are essential to successful simulation development. They enable a simulation development to use appropriate software development tools and techniques, to make good decisions about computer hardware and operating systems, to select an appropriate architecture, to choose appropriate software language(s) and environments, to produce appropriate documentation efficiently, to employ appropriate simulation and software development paradigms, etc. SMEs with expertise in simulations similar to the one to be developed have similar potential to help M&S development. They may provide useful guidance about algorithms, simulation structure, relationships among parts of the simulation, etc. Sometimes such SMEs are used informally for general guidance (occasionally even in brainstorming ways), and sometimes they are used in formal review processes. General insights in this paper about SME selection and management, about obtaining insights from SMEs and using them effectively in reviews, etc. apply as much in this arena as in any area of SME use.

---

[3] As a field that touches on many application domains -- ranging from Artificial Intelligence (AI) to software engineering, there are various definitions given for "knowledge elicitation," "knowledge acquisition," "knowledge engineering," "knowledge development," etc.

[4] Previous FDMS was called the Conceptual Model of the Mission Space (CMMS).

Simulation requirements are normally identified by the simulation sponsor or anticipated simulation users. Although such personnel can state what the simulation needs to be capable of doing, often these personnel are not expert in requirements engineering and may not produce a comprehensive, consistent, and cogent set of requirements which provide all the information needed to ensure that the simulation will satisfy its objectives without excessive guidance that unnecessarily restricts how the simulation can be designed and satisfy the requirements. Requirements validation SMEs help to ensure that the requirements are adequate, in an appropriate format, and fully represent intended user interests. Normally such SMEs are used in a review function, sometimes informally and sometimes formally (in which the review follows a specified process).

2B3. VV&A Expertise SMEs. It is important to understand possible relationships between VV&A SMEs and the VV&A team for a simulation. Some members of the VV&A team for a particular simulation may qualify as SMEs or technical experts because of technical expertise; such technical capabilities would not qualify that person for peer review since peer review (as noted in the descriptive definitions earlier) implies more independence than is likely for a member of the VV&A team[5]. Whether the SME or technical expert is a member of the VV&A team or an adjunct to it is immaterial, the principles and processes discussed in this paper apply to both equally. The administrative arrangements, including some of the SME selection processes, that are needed when SMES, technical experts, and those involved in peer review from outside the M&S team are obviously different than what can and normally is done when SMEs and technical experts come from within the M&S team.

We discuss four kinds of VV&A SMEs. First are those who have expertise in V&V techniques and methodologies. Next are SMEs who may be involved in a new kind of SME activity: producing estimates of simulation results that can serve as a surrogate for data in results validation when such real-world data are not available in acceptable quantity to be used directly. Third we discuss the traditional use of SMEs in qualitative simulation assessment, especially in "face validation." Finally we discuss SME usage in extending quantitative assessments.

2B3a. V&V Techniques and Technologies SMEs. There are many techniques and technologies that can be employed in M&S V&V. Most of the widely circulated lists of V&V techniques identify 40-50 techniques. This is true of a list by Balci [1994] which was also used in a 1998 M&S text [Pace 1998]. A much larger number of techniques exists; Balci [1998] and Binder [2000] identify more than 100 V&V techniques, and Miller et al [1993] identified more than 130 V&V techniques for software without addressing the specialized techniques employed with adaptive processing (such as Artificial Intelligence, also known as AI). An entire literature exists on verification, validation, and testing (and many techniques associated with such) of knowledge-based systems (KBSs), which includes expert systems [Plant nd; Preece et al 1995]. Likewise, there are many specialized techniques employed when formal methods are employed

[5] "VV&A team" is used very loosely since the composition of such teams vary widely (sometimes only one person, sometimes more in the "team"). Some M&S developments may not have a formal and distinct VV&A team; others may have formal identification of the lead for such, but leave all other participation in the team to ad hoc arrangements; and others may have formal identification of all involved in the various aspects of VV&A, and may even include separate arrangements for independent verification and validation (IV&V) of some or all of the M&S.

[NASA 1998; Murray 2002]. Many kinds of adaptive processes (traditional rule-based AI, KBS, fuzzy sets, neural nets, genetic algorithms, etc.) are now being used in computer simulations, and some employ formal methods, which means that the techniques and technologies of these fields now must also be considered in the complete kit of M&S V&V techniques and technologies.

Unfortunately, as discovered at a 1999 workshop on simulation validation [SIMVAL 99], ***most V&V/VV&A practitioners are not well-informed about many of the V&V techniques and lack the expertise needed for their effective use***. Many people involved in M&S V&V are unaware of current V&V literature; for example, many do not know that the Modeling and Simulation Information Analysis Center (MSIAC) published a survey of VV&A automated support tools in 2000 [MSIAC 2000]. Thus, SMEs with the technical expertise to employ specific V&V techniques can improve the quality of M&S V&V when such techniques are needed. Advice from VV&A SMEs about what techniques should be employed for a particular simulation is invaluable when the SMEs are competent in V&V technologies (understanding their real capabilities and limitations) and familiar with both the simulation application domain and technical characteristics of the M&S. Such advice can facilitate maximum correctness and credibility for the simulation given the resources and time that are available.

Although there are scores of V&V techniques, most group them into a relative small number of general categories. Balci [1994] uses six general categories: informal, static, dynamic, symbolic, constraint, and formal. Pace [1998] uses nine categories: review techniques, statistical techniques, techniques using logic and structure, techniques using sensitivity analysis, visualization techniques, M&S involving people, M&S with hardware in the loop, M&S involve AI and similar processes, and M&S requiring high integrity. Some techniques require specialized skills, such as the advanced mathematical training associated with lambda calculus, and may require specific software environments, as is the situation for many of the V&V techniques made possible by various computer-aided software engineering (CASE) tools.[6] Typically, based upon insights from those involved in software reuse, it takes at least three experiences with application of a technique to become proficient in it [Reifer 1998].

While the major part of this paper is focused on the role of SMEs in validation, their importance for efficient and effective verification should not be overlooked. Some communities, such as the computational science and engineering community, have so focused on validation issues that they have given inadequate attention to verification [Oberkampf and Trucano 2002], which is a major portion of what the software engineering community calls software quality assurance (SQA). Many do not understand the serious impact that lack of adequate verification and SQA can have on M&S.[7] V&V methods (techniques and technologies) SMEs can help a developer

[6] Cigital [2001] identified more than 2000 utilities and tools in the commercial market that provide computer aided support for software testing.

[7] From a seven year review of more than 100 mature programs of scientific software (i.e., programs in regular use which had been approved for "production use") that covered 40 application areas, including graphics, nuclear engineering, mechanical engineering, chemical engineering, civil engineering, communications, databases, medical systems, and aerospace – with both safety-critical and nonsafety-critical codes represented 8-12 serious static errors per 1000 lines of code were found in careful testing of the codes. The results of this review led its author to the conclusion that "the results of scientific calculations carried out by many software packages should be treated with the same measure of disbelief researchers have traditionally attached to the results of unconfirmed physical experiments." [Hatton 1997].

ensure that the best possible job of verification (supporting SQA) is done that can be done within time and resource constraints. Balci [2002 WSC] succinctly captures this idea: "An M&S application developed under an effective SQA program increases our confidence in accuracy much more than the M&S application developed without an SQA program." He later notes, SQA "is a profession, in and of itself, and having its own handbook" [Schulmeyer and McManus 1999].

2B3b. SME Opinion as a Surrogate for Quantitative Data.

Often SME opinions are used when quantitative data about the subject addressed by a simulation are limited or lacking. Typically SMEs will look at responses from the simulation or at the characteristics of a simulation (such as its algorithms) and then produce judgments about simulation capabilities or performance. This post-factum (*ex post facto*) approach is typical of "face validation" and other expert reviews. However, a key aspect of using SME opinion as a surrogate for quantitative data is collection of SME expectations prior to their exposure to simulation results. For this to be done in a meaningful way, it is very important to specify the scenario (or part of a scenario) that the simulation will address and to communicate that to the SMEs being used as the surrogate for quantitative data.. The collection of information given to the SMEs includes the full set of simulation inputs – both the factors that represent the system/subject whose behavior is being assessed and the conditions and circumstances that it will encounter in the simulation. Then the SMEs are asked to describe how the subject of the simulation will behave. Those expectations are collected and then used as data for comparison with simulation results.

When this is done, several things are important. First, SMEs must get adequate information to produce informed estimations of subject behavior and performance. Second, it is helpful to have a number of SMEs involved so that there will be enough "data" from their expectations to have confidence that one has obtained enough information for reasonable conclusions. Third, there is an inherent trade-off between preserving SME evaluation independence (so one opinion will not dominate or contaminate other opinions) and enabling SME synergy (so the whole is greater than the sum of the parts). Fourth, the SME estimations should be directly relatable to simulation outputs so that questions of interpretation and translation from one parameter to another are avoided. Finally, standard statistical processes can be used to determine coherence for SME estimations for various situations as well as for comparison of SME expectations with simulation results.

This process is new and is being experimented with to determine if it has enough merit and reliability so that it can become a standard "validation" approach for those simulations which address things that cannot be validated in a more objective and quantitative way. Metz and Harmon [2002] describe one of the more substantial experiments investigating this technique.

2B3c. Qualitative Assessment SMEs.

Qualitative validation has to be used when adequate acceptable real world data do not exist to permit quantitative validation and is based mainly upon SME and peer review. Qualitative validation is sometimes pejoratively identified as "subjective." That is unfortunate since quantitative validation also often contains subjective elements: the choice of some parameters for assessment and not others, selection of the level of agreement between simulation results, and real world data may be selected rather arbitrarily without clear and compelling relationship to functionality, etc. The term "subjective" should generally be avoided because of untoward connotations given it and because it is difficult to honestly determine which kinds of assessment should be labeled as subjective.

It should be noted that many M&S communities, especially those concerned primarily with M&S scientific applications such as computational fluid dynamics (CFD), do not like use of the term validation when real world data are not available. However, in the spirit of "all models are wrong, but some are useful,"[8] validation is a useful concept even when real world data are not available[9]. In such cases, *validation is the judgment process that declares the simulation technically is appropriate for its intended application (accreditation is the administrative or management decision that accepts that technical assessment).* The most common form of qualitative validation is called "face validation" which means that the M&S performs as expected in the opinion of selected SMEs – the term "face validation" goes back at least to the mid-1960s [Hermann 1967].

In the past, qualitative validation was often done without rigor, either in regard to SME selection, their evaluation criteria and processes, or other approaches used in the assessment processes. Consequently, the quality of assessments varied widely; and, in general, there was little capability for repeatability in qualitative validation assessment. This situation is very undesirable, and one would hope that it can be changed. Unfortunately, some mistakenly assume that such undesirable characteristics are intrinsic to qualitative validation.

Elaboration for qualitative validation should bring more rigor and standardization to SME selection, evaluation criteria, and assessment processes. A construct for elaboration for qualitative validation is presented below[10].

Meaningful qualitative validation is validation assessment whose significance is clearly understood and which has potential for consistency and possibly even for repeatability.

For meaningful qualitative validation:

> a *specified set* of *explicitly qualified SMEs* reviewing an M&S conclude that *specified characteristics* show *expected responses* for *specified portions of the mission space (i.e., the application domain)* for the M&S with expected responses identified.

---

[8] The origin of the quotation "all models are wrong, but some are useful" is usually attributed to statistician George E. P. Box, former president of the American Statistical Association. George E.P. Box, "Robustness is the Strategy of Scientific Model Building" in *Robustness in Statistics,* eds., R.L. Launer and G.N. Wilkinson, 1979, Academic Press, p. 202.

[9] Many confuse *fidelity* and *validity.* Fidelity is an absolute measure of simulation relationship to the reality represented (usually described in terms of resolution, accuracy, etc.). On the other hand, validity is a relative measure of whether the fidelity of a simulation is adequate to support its intended use.

[10] The following material is largely taken from Pace's paper on "Validation Elaboration" (presented at the 2002 Summer Computer Simulation Conference).

Each of the *italicized terms* is discussed below.

Explicitly qualified SMEs are SMEs who are qualified by (and their qualifications are documented): 1) Expertise, and/or 2) Vested interest.

A SME explicitly qualified by expertise is one who has standard expertise credentials such as pertinent formal education or training, specific relevant experience (such as involvement in similar assessments), publications related to the subject of the assessment, etc. In general, as suggested in the DoD Recommended Practices Guide (RPG) for M&S VV&A, qualifications of SMEs used in validation assessments should be documented and available to those making decisions about the simulation and its use. In situations in which use of the simulation has safety/critical implications or potential impact on major decisions or activities, it may be appropriate to follow guidelines used by courts for expert qualification: "factors the Supreme Court listed as relevant to the inquiry are: 1) whether the expert's technique can be tested through the scientific method; 2) whether the technique was subject to peer review and publication; 3) the known or potential rate of error of the technique; 4) the existence and maintenance of standards controlling the technique's operation; and 5) whether the technique had gained general acceptance in the relevant scientific community." [New Jersey Judiciary 2002].

A SME who is qualified by vested interest is one who belongs to an organization whose system or area of responsibility is addressed by the simulation, one who is potentially influenced by or influential relative to the simulation or its application, or one whose opinions about the simulation those responsible for the simulation believe is important. Normally it helps if there is an explicit documentation of reasons for selecting vested interest SMEs in a validation assessment.

> The **specified set of SMEs** used in a qualitative validation assessment should not only document why particular SMEs were selected for reasons of expertise and vested interests, but also indicate how SME evaluations are going to be used in the validation assessment: will a single SME evaluation be adequate for a conclusion, will a majority of SME evaluations be adequate for a conclusion, or will all SME evaluations have to agree for a conclusion?
>
> **Specified characteristics** of the simulation (parameters and combinations of parameters) that are the primary basis for the qualitative validation assessment should be explicitly identified *prior* to commencement of the assessment. Other characteristics of the simulation may also be used in the assessment, but not in a way that contradicts conclusions based on the specified characteristics.
>
> **Expected responses** of the simulation by the SMEs selected for the qualitative validation should be specified *prior* to commencement of the simulation assessment. Doing this is key to making qualitative assessment consistent and possibly even repeatable. Most qualitative validation assessments to date have not required SMEs to explicitly state their expectations prior to examination of simulation results; this is one of the main reasons that past qualitative validation assessments have not been consistent or repeatable.

The discipline of knowledge engineering has developed methods (some of which are discussed elsewhere in this paper) to elicit information and knowledge from domain experts in ways that facilitate meaningful and relatively consistent capture of such information; these methods can help ensure that SME assessments (both the *a priori* predictions of expected responses and the *a posteri* assessments of simulation results) are consistent and possibly even repeatable. Techniques are being developed to calibrate expectations from SMEs so that such can be used as a surrogate for real world data as a validation basis when real world data are not available; these techniques can also help bring consistency and possibly even repeatability in qualitative validation assessments [Metz and Harmon 2002].

> A meaningful qualitative validation assessment is focused on a **specified portion of the mission space** that the simulation addresses. It is widely recognized that SME review is most reliable for normal and nominal conditions, and that SME review is not very reliable for assessments of unusual or rare conditions [Pace 1999]. Specification before the assessment of the portion of the mission space that will be the primary basis of the qualitative validation assessment is important, and helps the assessment to have logical and factual soundness.

2B3d. SME Extension of Quantitative Assessment.

Quantitative validation can be used when validation can be based upon specific measurable aspects of M&S results, such as comparison of specific simulation parameters with real world data to determine the accuracy of representation of those parameters in the simulation. Obviously more confidence is likely to be placed in the correctness of simulation results when quantitative validation is possible than when one has only qualitative validation. However, for many simulations, a combination of qualitative and quantitative assessment processes will be needed to address the full scope of simulation capabilities[11].

Quantitative validation assessments, just as in the case of qualitative validation, have often been conducted in ad hoc and unsystematic ways so that one is not always certain what is meant by validation for a particular M&S even when quantitative approaches are used. Often validation for a particular simulation has been asserted on the basis of visual comparison of simulation parameter values on a chart which also shows real data[12]. While such may make for impressive presentation, it does little to significantly quantify the relationship of simulation results to real world data.

The question of how to meaningfully perform quantitative validation assessments is complex. It is necessary to determine the uncertainties associated with the real world data used as the basis of comparison, and then to specify exactly what kind of statistical comparisons are to be performed. This brings significant challenges for many in scientific and engineering communities. Often they do not have uncertainties for the real world data that they use appropriately characterized. Many of their computations (including simulations) are deterministic, and they do not normally

[11] The following material, as was done in the previous section, is largely taken from Pace's paper on "Validation Elaboration" (presented at the 2002 Summer Computer Simulation Conference).
[12] Various pejorative labels have been applied to this approach to validation.

use non-deterministic simulations (even though the reality they address has stochastic aspects). They may not have a good grasp of statistical issues that can be pertinent for such comparisons, and may have little appreciation for when it is more appropriate to employ classical statistical techniques and when Bayesian techniques are more appropriate. Those in other communities are not likely to be better off.

Fortunately, substantial attention has been given to this subject in the past few years, and some of these endeavors are mentioned here. The V&V guide for computational fluid dynamics (CFD) published in 1998 by the American Institute for Aeronautics and Astronautics (AIAA) focused attention on the need to explicitly and quantitatively address errors and uncertainties in real world data (experimental results) as well as in simulation results. The Department of Energy (DOE) Accelerated Strategic Computing Initiative (ASCI) program has sponsored significant effort in the area of validation metrics, which are part of validation quantification, and methodologies that should be associated with them [Trucano et al 2001]. Several of the DOE laboratories have started a major attack on the subject of uncertainty as it pertains to simulation results and their comparison with the real world [)berkampf et al 2000]. The Simulation Interoperability Standards Organization (SISO) Simulation Interoperability Workshop (SIW) has undertaken to improve understanding of what simulation fidelity is and how it can be measured [Gross et al 1999; Roza et al 2000]

Unfortunately, these endeavors have not yet resulted in widely-accepted standardized approaches that ensure quantitative validation assessments have clearly understood significance and that they are consistent and repeatable. The elaboration for quantitative validation suggested below can bring more rigor and standardization in such assessment processes.

For meaningful quantitative validation:

> for *specified data* about the real world (the system, process, or phenomena represented by the M&S), *specified parameters* available both from data and M&S results must satisfy *specified statistical relationships*.

Each of the *italicized terms* is discussed below.

> **Specified Data.** Data may come from experiment, test, observation, or specified subsets of such. All data have uncertainties associated with the data; unfortunately, such uncertainties may not be well characterized (identified, estimated, and documented). Such uncertainties include not only resolution and accuracy limits of measurements and observations, but also include less than complete and accurate descriptions of the conditions under which the data were collected. The later uncertainties can make it difficult to assure that simulation inputs correspond exactly to the conditions for which the data pertain. The **specified data** identify which data of that which might be pertinent are going to be used for comparison in quantitative validation assessment. Sometimes one may specify all the possible data are to be used. Other times one may specify only the data from a particular test (such as data from a particular missile test flight) or set of tests (such as only those tests conducted by a particular organization with a good

reputation for conducting tests well). Explicit identification of the data sources which will be used in the assessment is the first step in meaningful quantitative validation.

Specified Parameters. Modern complex simulations contain many parameters. For example, a model of the human heart [Buchanan 1999] that runs on Europe's largest supercomputer has 30 million equations! This kind of complexity makes it impossible to compare every simulation parameter with real world data. There just isn't enough time to do so, even if one had no other resource constraint. Thus, one must select **specific parameters** for comparison. There are two fundamental principles to guide that selection in quantitative validation assessments. First, data about a parameter selected must be available both from the simulation and from the real world. Second, the parameter should be a significant one from the perspective of intended use for the simulation.

Specified Statistical Relationships. When specific parameters from simulation results are compared with real world data, **specified statistical relationships** have to be demonstrated for quantitative validation assessment. Preferably, the evaluation criteria for the assessment (the particular statistical relationships that have to be satisfied) are specified before the assessment begins, not as a result of the assessment. It is much better to be able to say, "this is how good things must be for the simulation to be validated", than to say, "this is the best that we can do, so we will call that valid." Full-up operational tests of complex systems are chronically under sampled. Of the thousands (or millions) of possible scenario combinations, one or a few are actually measured. See Baker, et al "Comparison of MUVES-SQuASH with Bradley Fighting Vehicle Live Fire Test Results" for an example of the difficulty of comparing one (or a few) actual test outcomes to a manifestly complex set of interactions.

Specifying required statistical relationships (and then determining if such are satisfied) is not a trivial undertaking. Take a missile flight simulation and comparison of some of its parameters with data from a test flight. First, the test flight is one example of missile performance. A missile flight simulation typically employs a variety of distributed parameters to account for variations in manufacturing tolerances and other factors which can impact missile performance. Usually such a simulation is run a number of times, using monte carlo techniques to randomly vary such parameters. The initial question is, Should a single run be compared with the flight test data? Or, should the collection of runs be used? For example, should the mean of results for a particular parameter to be compared with the flight data? Or, should the comparison simply be that the data from the flight test not be outside the envelope of simulation results at any time? This very simple example illustrates some of the complexity of this subject.

The statistical criteria that are to be applied may also vary once the identity of the parameters is settled. A criterion may be as simple as the difference between the simulation parameter and the real world data may not exceed a specified value at any time, or the criterion may be a standard statistical test (such as acceptance of a hypothesis that the simulation data and the real world data are from the same data distribution at a specified confidence level).

Hypothesis testing has data implications. Data requirements for avoiding Type 1 Errors (rejecting good models) are different than data requirements for avoiding Type 2 Errors (accepting bad models). Availability of real world data may impact which statistic relationships are feasible. Many contend that data from a single test, regardless of how well it correlates with simulation results, provide an inadequate basis for validation conclusions. Fuller discussion of statistical process limitations relative to simulation validation may be found in Foundations ’02 papers by Berger, Easterling, and others.

Elaboration of quantitative validation must address the inference process when one wants to make a judgment about a simulation’s capability in a region for which real world data do not exist, as illustrated in the Figure 1.

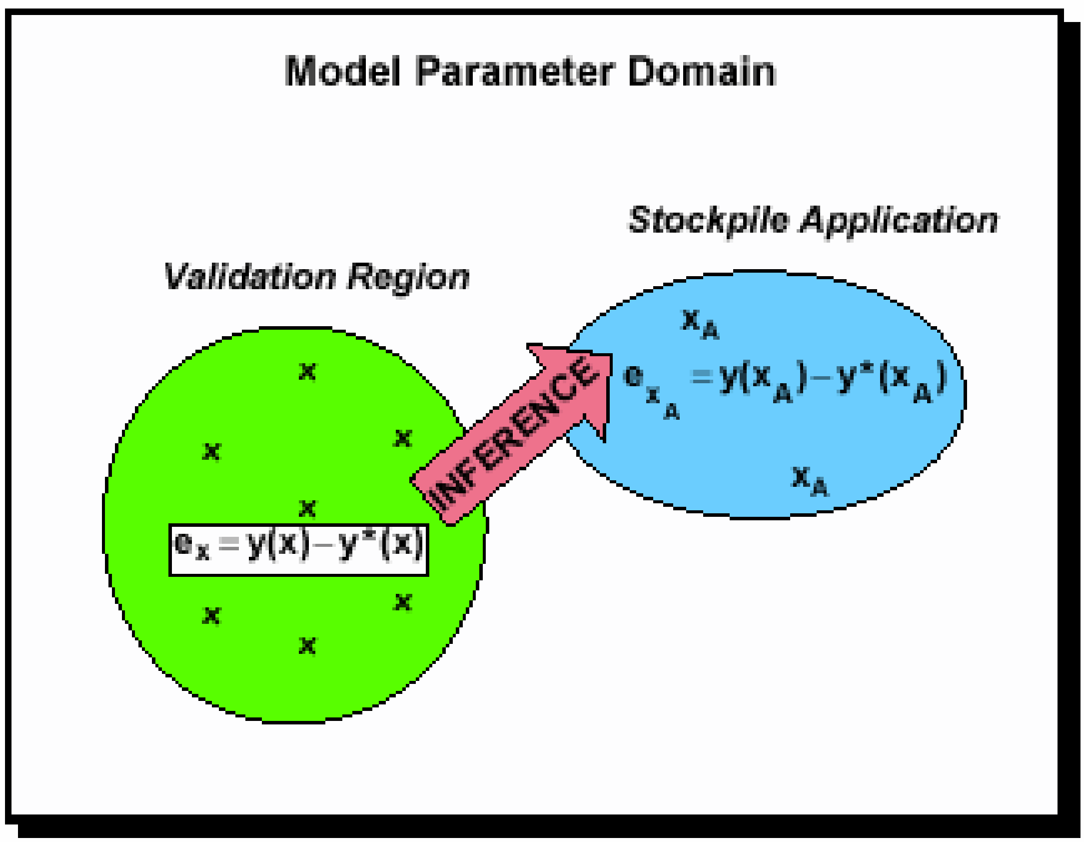

Figure 1 Extrapolation beyond the experimental validation regime (where real world data exist) is typically required for validation decisions about nuclear stockpile applications of computational science and engineering simulations. From R. G. Easterling, *Measuring the Predictive Capability of Computational Models: Principles and Methods, Issues and Illustration*s, Albuquerque, NM: Sandia National Laboratories Report SAND2001-0243, 2001.

The situation illustrated in Figure 1 is normally the case when validation reviews are conducted. If data exist for the conditions of interest, why bother to examine the subject by simulation. Just use the data. The problem that is usually faced has data for some conditions to which simulation results can be compared and one must estimate whether the simulation results will be appropriate for other conditions. Unfortunately, the processes for making that inference are frequently not made explicit in validation reviews. The rationale for such inference should be made explicit (and any evidence supporting it or casting doubt on it noted) for quantitative validation to have maximum meaning.

Some want to also consider results from accepted simulations as a form of "data." The processes suggested for quantitative validation elaboration above clearly apply in such a case, though it is probably better to reserve the term "data" for real world data and call results from simulations something other than data. The process of comparing simulation results with results from other simulations has been called "anchoring" in some communities when the simulation used for the comparison is one that has much greater fidelity than the one being reviewed for its validity. This is a process that is sometimes employed in military simulations when an engagement or campaign level simulations has its processes for things like missile fly-out anchored to 6-degree of freedom (DOF) simulations of the individual missiles. This process is typically used when there are relatively small amounts of real world data (only a few dozen actual flight tests or flights of the missile) and missile performance under many different conditions is required for the engagement or campaign level simulation[13].

2C. SME Activities within the SME Roles

2C1. Knowledge Production. The arenas in which SMEs are expected to produce knowledge related to model and simulation VV&A are very varied. They include knowledge about the application domain for the model or simulation, knowledge about processes and algorithms that will be used in the simulation to represent the characteristics and behavior of things represented by the simulation, knowledge about the real world that simulation responses are to represent, knowledge about computational equipment and processes that will be used in the simulation, and more. Comments in this section about production of such knowledge (elicitation of information from experts, transformation of that information into descriptions that are consistent across experts and with other information sources and that are amenable to automated processing, and preparation of information synopses that are easily consumed by humans) in general apply to all areas of SME usage in M&S VV&A.

Knowledge production is dominated by several objectives: 1) a disciplined procedure to systematically acquire knowledge about the real-world, tasks, and other simulation-related aspects that is sought from the experts, 2) a set of information standards that allow effective communication among the SMEs, information specialists, and others involved in the simulation's development and use, 3) methods for confirming that information specialists have

---

[13] Sometimes the magnitude of such endeavors is mind-boggling. In 1993 when the government made a choice (a "downselect") between two missile candidates for the one that would become the PATRIOT PAC-3 interceptor, nearly a million simulation runs were made for the analyses which supported that decision. Those simulation runs included 934,500 high-fidelity engineering-level missile fly-out simulation runs that defined intercept conditions so that a heuristic lethality model could determine consequences of the intercepts. That heuristic lethality model was anchored to both testing results (from sled tests and gas gun tests) and results from lethality hydrocodes. The heuristic lethality model was used because there was too little test data to simply use data to detemine intercept consequences from various intercept conditions and there was not time to run lethality hydrocodes for the many cases of interest.

correctly understood and represented information from SMEs, and methods to ensure that information provided by SMEs is correct, 4) methods for determining which information items are appropriate for possible re-use in other simulation situations, and 5) means (such as a library) for making information items which are potentially suitable for re-use readily available. Most of these objectives were addressed directly by DMSO's FDMS (previously CMMS) program [Sheehan et al, 1998]. DMSO created a set of knowledge development tools that addressed common semantics and syntax (CSS), a common format data base management system (DBMS), and digital interchange formats (DIFs) to support collection of authoritative knowledge about the military application domain for its FDMS library that was reasonably consistent across a wide variety of military operations. Others [e. g., Sharp] have focused on methods to ensure that the knowledge specialist has correctly captured information from the experts. The assessment of correctness of information provided by SMEs and other experts is the basic validation problem that one must always address.

2C2. Review of Plans-Designs-Products-Results. When SMEs are used in review, whether they review plans, designs, products (such as computer code), or simulation results, the specific foci of the reviews should be explicit: exactly what is expected from the SMEs should be clearly identified. Failure to make such expectations clear is a major source of problem in using SMEs, and the primary reason that recipients of SME reviews find them less useful than expected. Appropriate orientation for SMEs and use of a specified reporting format for SME reviews helps to avoid this problem and enhances the value of SME reviews. Use of a standard reporting format for SME reviews also helps to capture SME rationale for decisions and conclusions more completely and cogently than is likely otherwise.

Processes have been developed [Balci 2002, Metz & Harmon 2002] for calibrating SMEs and encouraging consistency from them. Employment of such processes enhances the value of SME reviews.

It should be noted that SME reviews are critically dependent upon the quality of the information provided them. If SMEs are reviewing plans (or products) and what is provided them is not the most current versions, then the reviews will be less valuable than they could have been. If the information provided SMEs is in a specialized format (such as the notation of a particular software development paradigm), the potential for misunderstanding of that information by the SME is high unless the SME is knowledgeable of that notation as well as the topic in which the SME is expert.

2C3. Advice. Use of SMEs for advice differs from use of SMEs for review. In SME reviews, the SMEs examine something that the simulation or its developer has produced (plans, designs, code, simulation results, etc.). SME assessments relate directly to the things reviewed. In giving advice, SMEs are asked to make judgment about things that are not defined as concretely. The two primary factors that need to be addressed in regard to SME advice are 1) the level of SME expertise, and 2) any factors that might influence (bias) SME judgment. An appropriate SME nomination form should provide the basic information needed for these factors to be considered appropriately – at least the information on the nomination form should provide pointers about where additional information would be helpful to insure that a SME's advice is appropriately understood.

2D. Selection of SMEs for Use within VV&A

Two primary considerations must be addressed in selecting SMEs for simulation VV&A. The first question is, "**Why?**" What is the SME's function? A partial listing of potential SME roles in simulation development was presented above. These imply some of the kinds of functions that SMEs might perform. SMEs can be used to provide timely, relevant, and credible testimony of first-hand knowledge or experience in the subject area of interest. This can result in reduced costs and development time.

The "Why?" question should be addressed explicitly before proceeding to the second question of **"Who?"** Can the SMEs needed be found within the simulation development team? Or must at least some of the SMEs be drawn from outside the simulation development team? In most simulation developments, many SME functions can be satisfied by members of the simulation development team. However, in most simulation developments, at least some SME functions can only be satisfied by those outside the simulation development team. Wisdom is needed for decisions about which functions can be addressed by simulation development personnel and which functions should be addressed by SMEs outside the simulation development team.

SMEs provide:

- personnel who really understand the subject (or parts of it) represented by the simulation (i. e., people who provide domain expertise),

- personnel familiar with the simulation (usually drawn from the simulation development team),

- personnel with appropriate simulation technical expertise (in the software, hardware, etc. expected to be used in the simulation, and in V&V techniques),

- personnel with background in similar simulations, and

- personnel with vested interests in the simulation (this community usually includes the simulation sponsor, its users, and those related to potentially competitive simulations).

If the validation review team includes representatives from all of these groups -- or at least reflects their interest, the topic is likely to be given more thorough validation review and thereby have more credibility for the simulation's applications.

2D1. Locating SMEs. Locating suitable SMEs with expertise in a particular subject is usually done 1) by going to those with whom the simulation sponsor, developer, or review authorities are familiar, or 2) by seeking recommendations from knowledgeable sources (such as the National Academy of Sciences, professional associations, experts in the field, etc.). Sometimes SMEs are asked to nominate themselves in response to announcements about the need for certain skills on the assessment team. Irrespective of which way one identifies a prospective SME, use of the kind of nomination form for SME candidates described below is likely to prove very helpful.

2D2. SME Nomination Forms. Many have found SME nomination forms useful in SME selection and management. Many organizations that use technical experts have developed forms related to their particular needs. Samples of such forms may be found on the internet under Peer Review Nomination, Expert Questionnaire, and similar titles. In general, these forms usually have several sections for the kinds of information indicated below:

- *SME Contact Information:* name, organization, position, address, phone/FAX, email, etc.;
- *SME Qualifications* (re potential areas of use): education, experience, positions, etc.;
- *SME Simulation Knowledge* (of the simulation in question and of simulation in general);
- *SME Availability*; and
- *Other Information* of possible pertinence.

The nomination form may also have sections for recording contact with the SME, decisions about the SME's suitability, use of the SME, etc. A spreadsheet or data base program can be very useful for maintaining such information about SMEs. Documentation about SME qualifications (whether positional, such as a representative of a vested interest like the program office developing a system represented by the simulation, or technical like the kind of qualifications indicated for the world class radar SME in one of the footnotes) can help to minimize criticism of SME reviews. Such documentation can also help a simulation sponsor and user develop a stable of SME candidates for use with a simulation at different points in the simulation's life cycle.

2D3. SME qualification

2D3a. Technical: Technical competence in the particular area for which a model or simulation is to be reviewed is usually the most important consideration in determining a SME's qualification. No formal processes for determining SME technical qualifications exist in any of the M&S VV&A policy and guidance documents known to the authors[14]. America's legal system has long wrestled with the issue of "expert witnesses". Their experience in this regard is useful for guiding how one addresses a SME's technical qualifications to participate in M&S VV&A reviews.[15]

---

[14] In disciplines which have been long established, there are explicit requirements for one to achieve various expert status (such as being licensed as a *Professional Engineer* or *Board Qualified* for various medical specialties). Even in relatively modern disciplines, there are very specific requirements for one to qualify as a pilot for a particular kind of aircraft. At this point, there are no professional status criteria that are widely accepted and used for credentials as a modeler, simulationist, analyst, or V&V practitioner – though efforts are underway to establish such criteria for simulation professional credentials: see the home page of the Modeling and Simulation Professional Certification Commission (M&PSCC), under the auspices of the National Training Systems Association (NTSA), which is responsible for development and provision of professional certification to simulation specialists (http://www.simprofessional.org/).

[15] It is important to remember the context of testimony in a court: generally it is to bring forth evidence about something, and when the evidence is such that the ordinary lay person might not easily understand it, to have it

A witness can be qualified as an expert by "knowledge, skill, experience, training, or education," but no particular requirements are specified. Evidence supporting such qualification can include academic degrees, time in the professions, pertinent accomplishments, etc. The more significant guidance from expert testimony for SME V&V review lies in the processes expected: testimony based upon "reliable facts or data," and testimony that is the product of "reliable principles and methods" which have been applied to the particular situation. What these processes suggest for SME V&V review are: 1) the basis for a SME's qualification should be explicit and available along with the SME's findings and conclusions (a SME nomination form, as discussed earlier, makes such information easily available), 2) the processes (techniques, methods) that a SME used in developing findings and conclusions should be specified and how those processes were applied should be described, and 3) caveats about the findings and conclusions should be very explicit (such caveats should include clear identification of parts of the SME's findings or conclusions which are opinions on topics that are outside the SME's area of expertise should any of the findings or conclusions be such).

In legal proceedings, the testimony of an expert witness may be challenged by the opposing counsel and subject to refutation or contradiction by another "expert." Few SME M&S VV&A reviews are subject to such critical scrutiny, even though the legal system has found this approach essential for determining the facts in many cases. This is one reason that it is prudent for V&V reviews of major simulations to include SMEs from different interest areas so that such challenges to findings and conclusions will occur within the V&V process (and possibly can be resolved at the working level instead of creating confusion about the simulation and its capabilities).

2D3b. SME Qualified by Vested Interest. Because perspective plays such an important role in judgments about M&S appropriateness for particular applications and the credibility that should be granted it, it is important that those with vested interest in the M&S or its applications be represented in review processes. Those responsible for a M&S, those likely to be significantly impacted by the simulation's use, and those able to influence the simulation's use have legitimate vested interest in the simulation and in ensuring that its results are properly understood. If the SME is to represent a vested interest such as a program office responsible for a system represented in a simulation, the best place to start to locate that kind of SME is the program manager for that system. This approach will normally identify a SME whom the program manager believes has appropriate technical competence and understanding of the program to play an important role in assessing the system's representation in the simulation. However, there are two important issues that often arise with such SMEs. One issue concerns their availability to participate in reviews of the simulation at the time desired, and the other

---

explained in ways that make it understandable. In this regard, "lay witnesses" in court are mainly restricted to stating facts. Admissible opinions from them are very restricted. On the other hand, "expert witnesses" are not only allowed to state facts (and explain them for lay consumption) but are also given much more latitude to express opinions related to their area of expertise. Much of the material in this section is drawn from Rule 701 (Opinion Testimony by Lay Witnesses), Rule 702 (Testimony by Experts), Rule 703 (Basis of Opinion Testimony of Experts), Rule 705 (Disclosure of Facts or Data Underlying Expert Opinion), and Rule 706 (Court Appointed Experts) in the *Federal Rules of Evidence* (available at http://www.house.gov/judiciary/evid2001.pdf) and various discussions of them (such as by Giannelli et al, 1999). Rules of evidence for states are generally similar to the federal rules of evidence issue concerns who pays for their participation in the reviews. Likewise, as a general rule, the primary place to start to locate SMEs to represent a vested interest is the office or organization with primary responsibility for that vested interest.

2D4. Personal Attributes Desired in SMEs.

SMEs involved in simulation VV&A activities require several attributes to be effective. (Note that this discussion uses the plural "SMEs" for situations that involve only one or more than one SME. In many situations, a single SME is adequate to accomplish the required function.)

Independence. SMEs must have adequate independence for honest and probing assessments. Great importance has long been attached to IV&V, in both software development [Lewis, 1992] and simulation [Williams, 1991]. It continues to be recognized within the software development community, e.g., the IEEE Standard for V&V [1998], and for M&S VV&A [Balci et al 2002]. The extent of independence required for a review team will vary with circumstances, but this factor should be addressed explicitly in planning simulation VV&A activities. Both real independence and the appearance of independence of team members are important. The first can impact simulation correctness, the second simulation credibility. It may be difficult to arrange convenient funding of "independent" members of the validation review team, i. e., people who do not belong to the M&S program manager, developer, or user organizations, unless those responsible for the simulation's development have created convenient mechanisms for such funding.

Recognized Competence. Competence is required to determine simulation correctness. Recognition of that competence can enhance simulation credibility. Competence requires that the total collection of SMEs on a V&V review team possess the knowledge and expertise required to perform the specified functions for which SMEs are needed. The review team typically needs a variety of expertise. Members of the simulation development team may possess some of this expertise, and some may be found only outside the team. Experience with simulations similar to the one being reviewed and experience with simulations of subjects similar to that represented by the simulation are also important for the review team. That background enables the review team to know where to expect problems. The team should be able to select a collection of V&V techniques and tools that will be capable of detecting both the most common kinds of simulation faults and the faults that have the greatest potential impact for damage to validity of the simulation's results. Use of SME nomination forms, such as described above, can help to ensure that appropriate recognition of SME competence occurs, especially if the information about SME qualifications is made available to those who have legitimate interest in knowing SME qualifications.

Trust of the Participants. The M&S program manager, user, developer, V&V team, and accreditation authorities need to trust and feel comfortable with the SMEs. The importance of the trust relationship in reviews has long been recognized in V&V literature [Pace 1986]. It helps if at least some of them know the individuals who will serve as SMEs, not just their organizations. Without confidence that an SME has no hidden agenda detrimental to the simulation development, the Developer is unlikely to "bare his soul" about the simulation's warts

[Glasow 1998]. Without knowing all of the potential problems of the simulation that the Developer knows, the SME cannot do a thorough assessment.

Good Judgment. SMEs must exhibit good judgment so that they can determine when the topic (requirements, simulation context, conceptual model, simulation results, etc.) has been sufficiently examined; ability to determine what is good enough is very important because exhaustive examination of a complex topic is not possible. Various approaches are employed to help SMEs exercise good judgment; for example, Balci [2002] presents an indicator-based methodology for structuring and integrating SME evaluations.

Perspective. SMEs must have the right objective. The purpose of a review team is to determine capabilities and limitations so that the simulation can be used appropriately and so that appropriate confidence can be placed in simulation results. This means that it must be evaluated relative to its intended use, not according to other criteria. Ensure that constructive objective always dominates a review team's efforts is a major challenge in most SME reviews.

2E. Management of SMEs Used in VV&A.

Efficient management of SMEs requires some kind of assignment and report tracking system. The sophistication needed for such a tracking system depends upon how many SMEs are involved, the size and importance of the simulation being assessed, and the importance of its application. (The more important the application, the more important comprehensive, formal tracking of SME assignments and reviews.) M&S requirements and acceptability criteria specify which simulation representations and capabilities require validation assessments. Therefore, the tracking system should make it easy to determine what capabilities have been reviewed and to promptly identify the reports related to the reviews and the conclusions of assessments. The tracking system should also allow monitoring of SME assignments; it should quickly show if some SMEs are not being used, if some are being used extensively, which assessments involve multiple SMEs, etc. Any modern database or spreadsheet package can be used for such a tracking system. However, whenever possible, the tracking system should be incorporated into the larger management process employed for the simulation. SME VV&A activities should be addressed in the same way that other elements of the simulation development and use are addressed (for scheduling, status, document control, etc.).

Effective use of SMEs requires appropriate orientation for them; careful attention to evaluation criteria, review processes, and report procedures; and diligence to keep SME use focused on what for the SMEs are being used for.

2E1. Orientation. SME orientation is essential for effective use of SMEs in simulation VV&A. SME orientation has four fundamental parts: general information, perspective, review process, and special topics.

2E1a. General information. This portion of SME orientation describes the simulation's purpose and provides information about its history or pedigree, who is developing it, who is expected to use it and how, how it is being developed (software and hardware considerations, development paradigm, and the like), the development timeline, etc. This kind of information helps the SME gain general understanding of the simulation.

2E1b. Perspective. In this part of the SME orientation, the SME is exposed to pertinent M&S requirements and acceptability criteria in order to gain an understanding of the specific standards of representational fidelity and functional capabilities against which the simulation is to be assessed. A common mistake SMEs make is to apply an inappropriate standard in review of a simulation. For example, a human-in-control simulation-based wargame normally may not need the same representational fidelity that is required of a high-fidelity system simulation supporting hardware-in-the-loop capabilities.

2E1c. Review Process. The third part of the SME orientation must address the review process. What kinds of information will be available to the SME? Both in terms of content and format. Will the SME have the full set of simulation requirements? The full statement of acceptability criteria? A complete description of the simulation conceptual model (or pertinent part of it if the SME is only concerned with the representation of a specific system or process within the simulation)? Simulation design documentation? Simulation code? An operational version of the simulation? Test results for the simulation? Documents used as sources for simulation algorithms and data? Results from related simulations? Results from past applications of the simulation? Opportunities for discussion with simulation development personnel and/or simulation users? Etc.

How is the review to be performed? Will the SME only review documents? Will the SME interact with simulation development and use personnel? Will the SME run the simulation? Etc.

How is the review to be reported? See the section below for a detailed discussion of this review reporting?

2E1d. Special Topics. The fourth part of the SME orientation will address any special topics required to facilitate the SME validation reviews. For example, if the simulation conceptual model is described significantly in simulation design format -- such as using Unified Modeling Language (UML) notation and constructs, or using one of the formal methods paradigms like Z++, it may be necessary to teach SMEs about the descriptive format to increase the likelihood that they will correctly understand the materials which they review. SME conclusions because of misunderstanding the simulation are not helpful to anyone involved in the validation assessment, and every effort should be extended to ensure that such do not occur.

2E1e. Orientation Methodology. Whether SME orientation is done with SMEs as a group, with individual SMEs, by merely providing each SME with an orientation document or by some more elaborate method will depend upon many factors and will have to be decided for each simulation and its associated evaluations. Availability of SMEs and availability of personnel to provide orientation are prime considerations.

2E2. SME Review Guidelines. Review guidelines should emphasize the importance of thorough documentation of the reviews. The goal of each verification or validation review is to provide enough evidence for a sound conclusion about the appropriateness of the simulation (or the part reviewed) for the purpose(s) specified. This requires the review guidelines to emphasize logical sufficiency (i. e., the review will produce adequate information to support a sound conclusion). This means that the critical issues must be identified and the data/information required to settle them must be specified. *It is helpful if the review guidelines can separate data and information from their interpretation so that disputes (if they should arise) about "facts" can be separated from the "significance" attached to interpretation of the facts.*

It is a good general practice to have all SMEs employ the same guidelines in V&V reviews for a particular simulation. This facilitates comparison of reviews by different SMEs for the same thing in the simulation as well as making it easier to assimilate the reviews of different parts of the simulation into a coherent whole.

Review guidelines should emphasize the importance of thorough documentation of all reviews.

2E3. SME Review Reporting. Report of a SME review should contain the following basic information:

- What is being reviewed – by name, version, date, etc. of the simulation or simulation element when such exist – and the purpose of the review (conceptual validation, results validation, requirements tracing to preliminary design, unit or integration test, etc.)

- Who participates (name, contact information, etc.)

- What information is used: documents, interaction with simulation development team members by name and date, etc.

- Scope and criteria for representational assessment and other evaluation criteria employed in the review

- Assumptions, algorithms, functional capabilities, tests, etc. explicitly related to the purpose of the review, addressed in appropriate detail to allow full understanding of bases for conclusions drawn by the review team

- Conclusion and synopsis of the review findings, clearly separating fact from interpretation, and explaining the significance of the findings

- Recommendations to improve simulation correctness or credibility, or the conceptual validation review process

Where possible, it is helpful to include an indication of the developer's attitude toward conclusions and recommendations presented in the review, with clear statements of the developer's rationale when there is disagreement with conclusions or recommendations of the review. Of special importance is an indication of whether such disagreements are related to the "facts" about the simulation or to interpretation regarding the significance of the facts.

2E4. SME Review Scheduling. Many practical considerations affect review scheduling. These include availability of information from the developer (e. g., conceptual validation reviews cannot be performed until the developer has completed description of the conceptual model for the simulation), availability of information to be used as the performance standard in results validation reviews, and availability of SMEs and other personnel. As a general rule, reviews should be scheduled at the earliest time that required information and personnel can support them efficiently.

2F. Common SME Problems and Concerns

2F1. Perspective, Performance, and Perception Problems. Perspective problems, performance problems, and perception problems are frequently encountered with SMEs. Each kind of problem is discussed in turn.

2F1a. Perspective Problems. Some SMEs have difficulty in assessing a simulation relative to its intended application. The SME may want to evaluate the simulation in some other context. For example, a SME may inappropriately fault the simulation for using approximations instead of more detailed and more accurate algorithms even though the approximations provide adequate accuracy for intended and expected simulation applications. Appropriate orientation for the SME can minimize this kind of problem, but it may also be necessary to take corrective action in managing SMEs should it become apparent that this problem has arisen. Typically a reminder to the SME is all that's required. Sometimes the review report should be revised so that it does not reflect an inappropriate perspective.

Sometimes SMEs have a particular agenda that they will pursue during their involvement in a validation review. The agenda may be overt, or it may be hidden. Every SME who represents a vested interest can be assumed to have an agenda of looking out for that interest during the review process. The potential problems arising from such agendas should be addressed in two ways. First, every SME assessment should strive to make the factual and logical bases of the assessment explicit and clear. This forces any "hidden" agenda to have a solid factual and logical foundation. Second, it is helpful to have a variety of perspectives (agendas) represented within the SME team so that no particular agenda can be pursued without challenge from the assessment of those with a contrary or different agenda.

2F1b. Performance Problems. Sometimes a SME will have trouble complying with the review and reporting schedule because of other demands on the SME. This problem can be avoided, or at least minimized, by realistic estimations of how long it should take a SME to perform a review and report on it coupled with reasonable scheduling plans based upon SME availability, required information availability, etc.

Sometimes SMEs will not follow specified review and reporting procedures. Typically this kind of problem can be resolved by a reminder to the SME about the procedure and, where it makes sense, modification of the procedure at the SME's suggestion.

Sometimes a SME will have difficulty understanding the simulation and may make an assessment based upon misconceptions. *Allowing the simulation developer an opportunity to respond to a preliminary version of the SME report provides an opportunity to correct such misconceptions prior to their becoming public and helps to ensure that the assessment is more complete than it might be otherwise*

2F1c. Perception Problems. Simulation developers, users of legacy simulations, and others with vested interest in the simulation's acceptance often cast a jaundiced eye on SMEs from outside their sphere and may have suspicions about SME competence, objectivity, etc. They will sometimes criticize a SME for not using appropriate perspective in the assessment, for not understanding the simulation correctly, for having a hidden agenda, etc. These kinds of problems can be ameliorated by use of a SME nomination form that explicitly documents a SME's qualifications, by use of a specified review and reporting process that emphasizes the facts and logic upon which an assessment is based, and by providing the simulation developer, user, etc. with an opportunity to respond to preliminary versions of SME reports and having a specified mechanism for such responses to become part of the final assessments (should the SME and the responder not come to a common view).

Another kind of potential perception problem arises when a SME has a special relationship to the simulation sponsor (or other significant decision maker). That SME's assessment may carry additional weight because of the trust which is invested in the SME [Viet 1996]. This kind of situation should be recognized, when it exists, and every effort made to ensure that the SME's assessment is factual and logically sound. This kind of situation can also be exploited legitimately by adroit selection of who briefs validation assessments.

2F2. Other Common SME Problems. Three kinds of common SME problems are discussed briefly in this section: 1) those pertaining to use of SMEs with distributed simulation, 2) those pertaining to use of SMEs with legacy (existing) simulations, and 3) those pertaining to use of SMEs with new simulations (or major modification to existing simulations). Most of the comments about SME perspective, performance, and perception problems discussed above also apply to all three kinds of problems discussed below.

2F2a. Distributed Simulation. V&V reviews related to a distributed simulation have some unique challenges. Most of the work to date on distributed simulation has focused upon interoperability standards so that technical capabilities for distributed simulation to function exist (sometimes called "technical interoperability"). The harder issues of how to discuss and how to measure or assess compatibility of the individual simulations (federates) in a distributed simulation (federation) have yet to be resolved (sometimes called "substantive interoperability"). V&V reviews are concerned primarily about such compatibility and about the appropriateness of a collection of simulations in a particular distributed simulation environment for addressing an application (i.e., V&V reviews should focus on substantive interoperability). At this time, rigorous methods do not exist for addressing these issues.[16] This makes it very important that SMEs participate in V&V reviews who have experience in such application so that at least past problems associated with such an application and with the individual simulations and their distributed simulation environment can be considered.

[16] The VV&A Forum of the Simulation Interoperability Standards Organization/Simulation Interoperability Workshop (SISO/SIW) is working to develop guidance that can become an IEEE standard for High Level Architecture (HLA) distributed simulations. A related standard was developed for consideration with Distributed Interactive Simulation (DIS) implementations, IEEE Standard 1278.4.

There are special V&V problems that may be associated with distributed simulations, especially large ones that involve actual systems (as in a simulation supporting a major military exercise). The actual components of the simulation may not be assembled until the simulation is run, which precludes testing the simulation to determine if it is performing correctly. SME reviews to provide judgment about caveats pertinent to simulation results and responses are very important, especially when human operators may control part of the simulation to force it to produce situations necessary for desired training or exploration of a topic. Some of the V&V challenges associated with such distributed simulations are described by Saunders [2002] in his paper about V&V for Millennium Challenge 2002.

2F2b. Legacy Simulations. Documentation of the conceptual model for many legacy simulations is limited or non-existent. Likewise, documentation of previous assessments of the simulation may be equally spotty. This requires either that substantial validation effort to be spent in re-engineering (developing) such information or that validation endeavors treat the simulation as a black box. There are significant logical limitations on both the level and scope of validation assessments that take the black box approach. In addition, a legacy simulation's reputation may color any validation assessment since a negative finding might call into question decisions and actions taken on the basis of previous results from the legacy simulation -- and even a positive finding about the simulation may not be able to impart a desired level of credibility to simulation results if parts of the community have negative impressions about the legacy simulation. This makes it important for the potential user of the simulation to have a clear understanding of the legacy simulation's reputation prior to initiation of validation reviews so that how the validation reviews are done (and by whom) can have a reasonable possibility of accomplishing their function.

2F2c. New Simulations. V&V reviews have the greatest potential with a new simulation or with a major modification to a legacy simulation. However, in order for the V&V reviews to be most effective and efficient, it is important that they be performed at appropriate times and in appropriate ways. Guidance suggested in this paper should help in use of SMEs in such reviews.

2F3. Guidance For Combining Inputs From Various SMEs. For small simulations (or for those for which V&V resources are very limited), only a single SME may be involved in simulation review. In which case, the guidance of this section is not needed. For larger simulations, and for those whose V&V resources allow more than one SME to be involved, guidance is needed for how to combine inputs from various SMEs – especially when the assessments by the SMEs contradict one another.

The first principle is to give priority (i. e., more consideration) to the assessment of a SME who is expert in the particular topic of the assessment than to other SMEs. For example, if two SMEs are reviewing representation of a military radar system in a simulation and one SME is a radar engineer and the other SME is a military officer with experience in using that kind of radar, whose assessment should be given priority would depend upon the specific aspect of the radar representation being addressed.

The second principle is to give priority to the SME assessment which has the more substantial factual and logical evidence in it. For example, if one SME showed that the derivation of an algorithm used in the simulation omitted a factor or had errors or low-fidelity approximations in it which could cause significant problems under some conditions and the other SME simply showed that the algorithm produced acceptable results for simple conditions, the first SME is the one whose assessment should be given the most weight.

The third principle is to make sure that SME assessments are actually addressing the same situation if some form of arbitration among them seems to be required. In many cases, what seems to be contradictory or inconsistent assessments by SMEs are actually supplementary assessments because the SMEs have focused on slightly different situations. In such cases, the SME assessments can simply be combined to greater a more complete assessment of the simulation.

The fourth principle is to make sure that all SME assessment perspectives are included in V&V reports about the simulation, and explain why more emphasis is placed upon some SME assessments than on those of others. This kind of candor can be embarrassing at times, but it always enhances simulation credibility and confidence in its capabilities when it is clear that all assessments were given due consideration.

Finally, when SME assessments agree, it adds credibility to their conclusions – especially if the SMEs come from different communities. Credibility of the simulation based upon SME assessments can be enhanced if the diverse background of SMEs is described along with their conclusions about the simulation.

3. Major Issues

Major issues with SMEs fall into three main areas: 1) issues related to knowledge development (elicitation, acquisition, etc.), 2) SME selection and qualification, and 3) quality of review processes.

The issues related to knowledge development are concerned with how the knowledge produced for the simulation (whether domain knowledge that guides simulation development or knowledge used in evaluation of simulation capabilities) is judged for correctness, appropriateness, etc. We have not yet matured in this arena to the place where guidance that produces step by step procedures that assure success can be written. This increases the importance of using SMEs who have requisite experience and expertise in knowledge development along with SMEs with the pertinent domain knowledge.

The main SME selection and qualification issue is fundamental to M&S V&V: how to relate V&V review of specific M&S characteristics to ensure that the simulation has capabilities to satisfy its requirements and the implications for the kinds of SMEs needed to make that happen.

The final issue area, quality of review processes, is how much formality is necessary to prevent inappropriate conclusions from SME reviews and to ensure that appropriate SME review findings have the needed impact on simulation credibility. The guidance provided in this paper about that subject should prove very useful.

4. Major Research Areas Required For Significant V&V Progress

Research suggestions in this section come from three primary sources: 1) past suggestions about VV&A research that have not yet been accomplished, 2) current research needs recognized by the knowledge engineering community that are pertinent to SMEs, and 3) insights from the authors of this paper (including insights obtained from advanced contact with those involved in the Foundations '02 Special Interest Topic on *V&V Research*).

4A. Past Suggestions of VV&A Research

Sargent et al [2000] and Balci et al [2002] contain recent suggestions for VV&A research. We are not aware of significant responses to suggestions in these two sources, and treat their suggestions as still being pertinent because these needs have not yet been addressed. The suggestions in them which pertain to SMEs are:

- Research about how simulation V&V should change with M&S size, type, and application.
- Research to develop better cost estimation processes for M&S V&V, and how to separate SME review costs from other V&V costs.
- Research about how to use visualization capabilities to enhance SME reviews.
- Research about better ways to connect statistical processes appropriately to SME validation reviews.
- Research for how to better disseminate insights from VV&A experiences to M&S/V&V communities.
- Research for better ways to provide computer support (i. e., automation) for VV&A, especially for validation and accreditation. Balci [to appear 2002] presents an indicator-based methodology for structuring and integrating SME evaluations that illustrates this kind of automation.

The following suggestions by Feinberg et al [2001], drawn from their survey of automated VV&A tools, are pertinent to SME VV&A research needs, and in some cases repeat research needs from above.

- *Develop more automated support tools for VV&A.* The M&S community should emphasize developing new automated testing tools to support effective VV&A. There is a special need for tools supporting the verification and validation of distributed simulations and even more so for high fidelity distributed simulations. These tools should be analogs of the automated testing tools used in the testing and evaluation community, and analyze all the details of model interactions during exercise runtime.

- *Adopt or adapt tools from the software industry.* The M&S community should adopt or adapt tools from the software industry to support verification and validation as possible. The software industry provides many tools already used frequently in VV&A efforts. Software tool providers are well supported in their industry and are developing new tools based on current technologies as well as entirely new approaches to verifying and validating software code. These developments should be monitored for potential applications to M&S.

- *Make better use of visualization tools.* The M&S community should make more and better use of visualization tools in support of VV&A. Visualization tools help M&S developers and users to determine if a system "looks right" and if it is representing reality correctly. These tools support understanding of data, results, and system dynamics. Tools such as CAD/CAM viewers, 3D walkthroughs, exercise stealth viewers, and graphing packages for statistical analysis are all directly applicable to VV&A.

4B. SME-related Knowledge Engineering Research Needs

Effective techniques are needed for SME knowledge elicitation, representation, and integration. Ford and Sterman [1997] indicate that "increased clarity and specificity are required concerning the methods used to elicit expert knowledge for modeling." Rush and Wallace [1997] state that "this process proves most difficult when the elicitation and representation of knowledge from multiple experts is necessary." We believe that these areas are likely to be pursued within the knowledge engineering arena, without emphasis on their importance for SME use in M&S VV&A; but it is important that insights from such research be disseminated about simulation V&V personnel so that they can be used effectively.

4C. SME V&V Research Suggestions From This Paper's Authors

We endorse the research needs identified in the two previous sections and identify several other research needs to improve SME use in V&V. First, we believe that additional research is needed to identify processes which can bring more consistency in SME assessments and which enable identification of SMEs whose assessments are excessively influenced by bias and other factors besides simulation capabilities and characteristics. Second, we believe that additional research is needed to identify processes which ensure that SME knowledge is correctly captured by any automated or formal notations/formats and which provide awareness that such correct knowledge capture and representation has occurred. Third, we believe that additional research is needed to determine what characteristics of an expert truly qualify that person to perform expert validation reviews of simulations.

5. Recommendations and Suggestions

SME reviews are crucial for establishing M&S correctness and credibility for many simulations. We recommend that the suggestions of this paper about processes that can improve the quality and consistency of SME reviews be implemented widely, and made standard processes by those who control M&S development and use. This means more formality in processes used to determine SME qualifications, more formality in the ways that SMEs reviews are conducted and reported, and more formality in the ways that SMEs are oriented for their tasks and managed. We also recommend that the research needs identified above be addressed.

Our recommendations are addressed both to M&S V&V professionals and to those who fund, develop, and/or use M&S. Neither group alone can adequately implement our suggestions; action from both is needed if there is to be substantial improvement in SME use for M&S V&V review.

6. Conclusions

This paper has addressed the important subject of SME V&V review. For many, perhaps most M&S, SME reviews will play a major role in M&S V&V, perhaps the dominant role in many situations. We believe that SME reviews should play important roles, even in M&S V&V efforts which are able to produce quantitative validation results because there will be areas (in most cases) which can not be addressed simply by quantitative means, and SME reviews can be very helpful in them. Therefore, it is important that SME reviews be done in a way that makes them most useful in establishing simulation correctness and credibility. The ideas presented in this paper can help that happen, and we encourage that these ideas be used by all.

While we are confident that the ideas we present will help M&S V&V to be better, we realize that we do not have all the answers. The research needs identified in this paper, when addressed, will provide additional information that will enable SME V&V reviews to be even more effective, and we urge that those research needs be pursued vigorously.

7. Acknowledgments

Support from DMSO enabled the authors to prepare this paper and participate in Foundations '02; however, the views expressed in the paper are those of the authors and do not necessarily represent the views of DMSO or any other organization (public or private). We have drawn freely from the experiences and insights of many who have used experts in M&S V&V. They are too numerous to identify individually, and if we tried to list them all we would assuredly omit some; therefore, we chose to acknowledge them in this way. Most of these insights were given to us verbally (we have tried to identify those in print through our references to the sources identified in the following section); some insights we obtained simply by observing how experts were used in M&S V&V. The sources in the following section identify all the published material that we drew upon for this paper.

8. Bibliography/References

American Institute for Aeronautics and Astronautics (AIAA) Guide G-077-1998, *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*, January 14, 1998.

William E. Baker, et al, "Comparision of MUVES-SQuASH with Bradley Fighting Vehicle Live-Fire Results," Army Research Laboratory, Aberdeen Proving Grounds, ARL-TR-1846, November 1998.

Osman Balci, "Validation, Verification, and Testing Techniques Throughout the Life-Cycle of a Simulation Study," *Annals of Operations Research*, Volume 53:121-173, 1994.

Osman Balci, "Verification, Validation, and Testing", In *The Handbook of Simulatio*n, J. Banks, Ed., John Wiley & Sons, New York, NY, Aug., pp. 335-393, 1998.

Osman Balci, Richard E. Nance, and James D. Arthur, "Expanding Our Horizons in Verification, Validation, and Accreditation Research and Practice, **To be published in** *Proceedings of the 2002 Winter Simulation Conferenc*e, San Diego, CA, Dec. 8-11, 2002 (http://manta.cs.vt.edu/balci/papers/VVAHorizons.pdf).

Osman Balci, "A Methodology for Certification of Modeling and Simulation Applications," *ACM Transactions on Modeling and Computer Simulatio*n, to appear 2002 http://www.cs.dartmouth.edu/~tomacs/upcoming/11.4balci.pdf.

R. V. Binder, R.V., *Testing Object-Oriented Systems: Models, Patterns, and Tools*. Addison-Wesley, Reading, MA, 1191 pp, 2000.

Mark Buchanan, "The Heart That Just Won't Die," *New Scientist*, March 20, 1999, http://archive.newscientist.com/archive.jsp?id=21785100.

Cigital, "Testing Tools Supplier List," Cigital, Dulles, VA. (http://www.cigital.com/marick/faqs/tools.htm), 2001.

Department of Defense (DoD) *Recommended Practices Guide (RPG) for Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A)*, Millennium edition, 2000.

Jerry M. Feinberg, Patrick W. Goalwin, Pamela L. Mayne, and Phillip Abold, "Verification, Validation, and Accreditation (VV&A) Automated Support Tools: State of the Art Report - A Summary," *Proceedings of the 2001 Summer Computer Simulation Conference,* CD.

D. N. Ford and J.D. Sterman , "Expert Knowledge Elicitation to Improve Mental and Formal Models," *System Dynamics Review 1*4, 4, 309-340, 1997, http://web.mit.edu/jsterman/www/ford_sterman_elicit_1.html .

Marc S. Friedman & Stanley H. Kremen, "Technical Experts in Computer-Related Performance Litigation," *Computer Forensics Online*, Vol. 1, No. 1, Dec. 1997, http://www.shk-dplc.com/cfo/issue%201/expert.html.

Priscilla A. Glasow, "Characteristics, Selection, and Use of Subject Matter Experts," Technical Report 99W0000048, POET WG2K IV&V Paper No. 25, The MITRE Corporation, McLean, VA, June 1998.

D. C. Gross, et. al., "Report from the Fidelity Implementation Study Group," *1999 Spring Simulation Interoperability Workshop*, March 1999, Paper 167.

Paul C. Giannelli, Albert J. Weatherhead III, & Richard W. Weatherhead, Expert Qualifications & Testimony, **Conference on Science and Law,** San Diego, CA, April 15, 1999, http://www.scientific.org/distribution/law-review/giannelli.pdf.

L. Hatton, "The T Experiments: Errors in Scientific Software," *IEEE Computational Science & Engineering*, Vol. 4, No. 2, pages 27-38, 1997.

C. F. Herman, "Validation problems in games and simulations with special reference to models of international politics" *Behavioral Science, 12* (3), 1967, pp. 216–231.

IEEE Standards for Software Verification and Validation, IEEE Standard 1012, IEEE Computer Society, Washington, DC 1998.

Robert O.Lewis, *Independent Verification and Validation: A Life Cycle Engineering Process for Quality Software*. John Wiley & Sons, N.Y., 1992.

Michael Metz and Scott Harmon, "Using Subject Matter Experts for Results Validation - A Progress Report, paper 02S-SIW-095, Spring 2002 Simulation Interoperability Workshop (SIW), http://www.sisostds.org/conference/View_Public_Number.cfm?Phase_ID=2.

L. A. Miller, E. Groundwater, and S. M. Kirsky, *Survey and Assessment of Conventional Software Verification and Validation Methods,* U.S. Nuclear Regulatory Commission Report NUREG/CR-6018, 1993.

Modeling and Simulation Information Analysis Center (MSIAC), *Verification, Validation, and Accreditation (VV&A) Automated Support Tools: A State of the Art Report*, 2000.

Lynee D. Murray, "A Brief Overview of Formal Methods for Verification &Validation," *Proceedings of the 2002 Advanced Simulation Technologies Conference (ASTC), April 14-18, 2002*, Society for Computer Simulation International, 2002.

NASA, *Formal Methods Specification and Verification Guidebook for Software and Computer Systems: Volume I, Planning and Technology Insertion & Volume II, A Practitioner's Companion*, 1998.

New Jersey Judiciary: *njcourtsonline.com*, "Notice to the Bar: 2000 – 2002 Report of the Supreme Court Committee on the Rules of Evidence," *http://www.judiciary.state.nj.us/reports/indexsupp.htm*, February 8, 2002.

William L. Oberkampf, Sharon M. DeLand, Brian M. Rutherford, Kathleen V. Diegert, and Kenneth F. Alvin, *Estimation of Total Uncertainty in Modeling and Simulation,* Sandia National Laboratories Report SAND2000-0824, April 2000.

William L Oberkampf and Timothy G. Trucano, *Verification and Validation in Computational Fluid Dynamics*, Sandia National Laboratories Report SAND2002-0529, March 2002.

Dale K. Pace, "A Perspective on Simulation Validation," *Proceedings of 1986 Summer Computer Simulation Conference*, July 1986, pp. 187-189.

Dale K. Pace, "Verification, Validation, and Accreditation (VV&A)," Chapter 11 in *Applied Modeling and Simulation: An Integrated Approach to Development and Operation*, Cloud D.J. and L.B. Raines (eds.), McGraw-Hill, 1998, pp. 369-410.

Dale K. Pace, "Use of Subject Matter Experts (SMEs) in Simulation Evaluation," paper 99F-SIW-018, Fall 1999 Simulation Interoperability Workshop (SIW), http://www.sisostds.org/doclib/doclib.cfm?SISO_FID_3321.

Dale K. Pace and Priscilla A. Glasow , "SIMVAL 99 Final Report," Military Operations Research Society (MORS), March 26, 1999 (http://www.mors.org/Pubs/simval99.pdf). Overview and synopsis of findings in, "SIMVAL 99," *PHALANX*, Vol 32., No. 1, March 1999, pp. 22-25.

Dale K. Pace, "Validation Elaboration," *Proceedings of the 2002 Summer Computer Simulation Conference,* San Diego (CA), July 14-18, 2002.

Peer Reviews, SW-CMM- V2.0 Draft C, Software Engineering Institute (SEI), October 20, 1997, http://www.sei.cmu.edu/cmm/draft-c/c37prv.html.

Robert T. Plant, *Tools for the Validation & Verification of Knowledge-Based Systems: 1985-1995 References*, no date (http://www.csd.abdn.ac.uk/~apreece/Research/vvtools.html; accessed 7 April 2002).

Alun Preece, Pedro Meseguer, and Bob O'Keefe, *Verification, Validation & Testing of Knowledge-Based Systems: An Annotated Selective Bibliography,* 1995 (http://www.csd.abdn.ac.uk/~apreece/Research/vvbiblio.html; accessed 7 Apr 02 -- last updated in 1995).

Donald J. Reifer, *Practical Software Reuse*, New York: Wiley Computer Publishing, 1998.

Peter Roberts, "Scholarly Publishing, Peer Review and the Internet," *First Monday: Peer-Reviewed Journal on the Internet*, 1999.

Z. C. Roza, D. C. Gross, and S. Y. Harmon, "Report Out of the Fidelity Experimentation ISG," *2000 Spring Simulation Interoperability Workshop*, March 2000, Paper 151.

R. Rush and W.A. Wallace (1997), "Elicitation of Knowledge from Multiple Experts Using Network Inference," *IEEE Transactions on Knowledge and Data Engineering* 9, 5, 688-696.

R. G. Sargent, P.A. Glasow, J.P.C. Kleijnen, A.M. Law, I. McGregor, and S. Youngblood, "Panel: Strategic Directions in Verification, Validation, and Accreditation Research," In *Proceedings of the 2000 Winter Simulation Conference* (Dec. 10-13, Orlando, FL), J.A. Joines, R.R. Barton, K.Kang, and P.A. Fishwick, Eds., IEEE, Piscataway, NJ, pp. 909-916, 2000.

Randy Saunders, "Approaches to V&V on Millennium Challenge 2002," paper 02F-SIW-051, *Proceedings of the Fall 2002 Simulation Interoperability Workshop (SIW), available at http://www.sisostds.org/conference/View_Public_Tracks_Detail.cfm?Track_ID=19&Phase_ID=2*.

G. G. Schulmeyer and J.I. McManus, Eds., *Handbook of Software Quality Assuranc*e, 3rd ed. Prentice-Hall, Upper Saddle River, NJ, 1999.

John K. Sharp, "Validating an Object Oriented Model," *Journal of Conceptual Modeling*, Issue 5, December 1998, http://www.inconcept.com/jcm/ (this and other articles by Sharp on NLM can be found at http://www.sharp-informatics.com/index2.htm).

Jack Sheehan, LTC Terry Prosser, Maj Harry Conley, Lt Kevin Yentz, and Janet Morrow, "Conceptual Models of the Mission Space (CMMS): Basic Concepts, Advanced Techniques, and Pragmatic Examples," paper 98S-SIW-127, *Proceedings of the Spring 1998 Simulation Interoperability Workshop (SIW)*, available at http://www.sisostds.org/doclib/doclib.cfm?siso_FID_1180.

Timothy G. Trucano, Robert G. Easterling, Kevin J. Dowding, Thomas L. Paez, Angel Urbina, Vicente J. Romero, Brian M. Rutherford, and Richard G. Hills, *Description of the Sandia Validation Metrics Project*, Sandia National Laboratories Report SAND2001-1339, August 2001.

C. T. Veit, "Judgments in Military Research: The Elusive Internal Validity Issue," *PHALAN*X, 29(3), 1996.

"Verification, Validation, and Accreditation (VV&A) Automated Support Tools **-** A State of the Art Report Part 1." Modeling and Simulation Information Analysis Center (MSIAC), December 15, 2000.

Marion Williams, "SIMVAL Mini-symposium – A Report," *PHALANX* (June) 24(2): 1, 3-6, 1991.